Nr. 18
11. January 2015

Leopold-Franzens-Universität Innsbruck

Preprint-Series: Department of Mathematics - Applied Mathematics

Aggregation of Regularized Solutions from Multiple Observation Models

Jieyang Chen, Sergiy Pereverzyev Jr., Yuesheng Xu

Technikerstraße 13 - 6020 Innsbruck - Austria
Tel.: +43 512 507 53803 Fax: +43 512 507 53898
https://applied-math.uibk.ac.at

Aggregation of Regularized Solutions from Multiple Observation Models

Jieyang Chen[*], Sergiy Pereverzyev Jr.[†], and Yuesheng Xu[*‡]

January 11, 2015

Abstract

Joint inversion of multiple observation models has important applications in many disciplines including geoscience, image processing and computational biology. One of the methodologies for joint inversion of ill-posed observation equations naturally leads to the multi-parameter regularization, which has been intensively studied for several years. However, problems such as a choice of multiple regularization parameters remain open. In the present study, we discuss a rather general approach to the regularization of multiple observation models, based on the idea of the linear aggregation of approximations corresponding to different values of the regularization parameters. We show how the well-known linear functional strategy can be used for such an aggregation and prove that the error of a constructive aggregator differs from the ideal error value by a quantity of a order higher than the best guaranteed accuracy from the most trustable observation model. The theoretical analysis is illustrated by numerical experiments with simulated data.

Keywords: Multi-parameter regularization, joint inversion, linear functional strategy, aggregation.

1 Introduction

In various application fields, especially in geoscience, one is often provided with several data sets of indirect observations of the same quantity of interest, where each set contains the data measured by different physical principles. For example, when determining the gravity field of the Earth by satellite data we have to combine different types of present or future data such as GPS observations for the satellite missions CHAMP [31] or GRACE [32], with its satellite-to-satellite tracking (SST), as well as gradiometer measurements for the satellite mission GOCE [30], with its satellite gravity gradiometer (SGG). Other examples are related to various problems in geomagnetism, which require a combination of satellite data and ground data at the Earth's surface in order to obtain high resolution models (see, e.g., [10]). There is an extensive literature where SST- and SGG-problems were studied and treated separately and independently. For this subject, the readers are referred to the monograph [9] and to the references cited therein.

With multiple observation models, such as SST and SGG, each providing approximations of the same quantity of interest, one is left with the choice of which to trust. A more advanced question is what to do with a less trustable approximation, or what is the same, whether the approximation that involves all available observations may actually serve as an effective way to reduce uncertainties in independently inverted models. This question was and is discussed quite intensively in the geophysical literature, where the term "joint inversion" was introduced by the authors of [29] for the methods which give the solution of various types of observation equations inverted simultaneously. Similar ideas were also applied in medical applications, e.g., the joint utilization of CT and SPECT observations to improve the accuracy of imaging [4]. A short overview about the application of joint inversion in geophysics may

[*]Guangdong Province Key Lab of Computational Science, School of Mathematics and Computational Science, Sun Yat-sen University, Guangzhou 510275, P. R. China. *chenjieyang3@126.com*.

[†]Institute of Mathematics, University of Innsbruck, Technikerstrasse 13, 6020 Innsbruck, Austria. *sergiy.pereverzyev@uibk.ac.at*.

[‡]*yxu06@syr.edu*. All correspondence sent to this author, who is also a professor emeritus of mathematics at Syracuse University, Syracuse, New York 13244, USA.

be found in [12], where it was mentioned in particular that there was no standard standpoints using the appellation joint inversion. To distinguish inversion methods based on different data combinations, researchers have introduced different names, e.g., aggregation [26], which was also used in the context of statistical regression analysis [15, 16]. Regardless of their names, what is in common for the above mentioned approaches is that they induces stability by simultaneously utilizing different types of indirect observations of the same phenomenon that essentially limits the size of the class of possible solutions [3].

The idea of the joint inversion naturally leads to the methodology of the multi-parameter regularization, which in the present context may be seen as a tool for the compensation of different observations. Multi-parameter regularization of geopotential determination from different types of satellite observations was discussed in [17]. Similar approach was adopted in the high resolution image processing and displayed promising effects in experiments [22]. At this point it is important to note that one should distinguish between multi-parameter schemes, where the regularization parameters penalize the norms of the approximant in different spaces, and the schemes, where the parameters weight the data misfits in different observation spaces. In the former schemes an observation space is fixed, and by changing the regularization parameters we try to find a suitable norm for the solution space, while in the latter schemes the situation is opposite: by changing the parameters we try to construct a common observation space as a weighted direct sum of given spaces. The choice of the regularization parameters for the former schemes has been extensively discussed in the literature. A few selected references are [6, 7, 8, 14, 21]. As to the latter schemes (schemes with a fixed solution space), we can indicate only the paper [17], where a heuristic paremeter choice rule is discussed, and the paper [18], where the parameter choice is considered as a learning problem under the assumption that for similar inverse problems a suitable parameter choice has been known. It is clear that such approaches can be used only for particular classes of problems.

Although a simultaneous joint inversion by means of multi-parameter regularization provides acceptable solutions, it still faces methodological difficulties such as the choice of regularization parameters that determine suitable relative weighting between different observations. The goal of the present study is to discuss a rather general approach to the regularization of multiple observation models, which is based on the idea of a linear aggregation of approximations corresponding to different value of the regularization parameters.

This paper is organized as follows. In Section 2 we discuss an analog of the Tikhonov-Phillips regularization for multiple observation models. In Section 3 we study a linear aggregation of the regularized approximate solutions and its relation to the linear functional strategies [1, 2, 19]. In Section 4 we illustrate our theoretical results by numerical experiments. We draw conclusions in the last section.

2 Tikhonov-Phillips Joint Regularization

In this section, we analyze the methodology from multiple observations to the joint regularization, and lead to the multi-parameter regularization in general form with the SST- and SGG-problems as an incident. Then, different parameter choice schemes in literature are investigated and which arises the topic about utilizing the diversity of parameter choices. We begin with setting up a general framework.

If m different kinds of observations are assumed, then in an abstract form they enter into determination of the quantity of interest $x = x^\dagger$ (e.g., the potential of the gravity or magnetic field), leading to the observation equations

$$y_i^{e_i} = A_i x + e_i, \ i \in \mathbf{N}_m, \tag{1}$$

where A_i denotes the design operator, which can be assumed as being a linear compact and injective operator from the solution space $\mathcal{X}$ into the observation space $\mathcal{Y}_i$, and e_i denotes the error of the observations, and $\mathbf{N}_m := \{1, 2, \dots, m\}$.

In a classical deterministic Hilbert space setting, $\mathcal{X}$ and $\mathcal{Y}_i$ are assumed to be Hilbert spaces and

$$\|A_i x^\dagger - y_i^{e_i}\|_{\mathcal{Y}_i} \leqslant \varepsilon_i, \ i \in \mathbf{N}_m, \tag{2}$$

where $\varepsilon_i \in (0, 1)$ is a noise level. In particular, for the spherical framework of the SST-type or the SGG-type problem, the design operators in model (1) have the form

$$A_i x(t) := \int_{\Omega_{R_0}} h_i(t, \tau) x(\tau) \, \mathrm{d}\Omega_{R_0}(\tau), \ t \in \Omega_{\rho_i}, \quad i = 1, 2, \tag{3}$$

where in case of SST

$$h_1(t,\tau) := -(4\pi R_0)^{-1}\frac{\partial}{\partial\rho}\left[\frac{\rho^2 - R_0^2}{(\rho^2 + R_0^2 - 2\,\langle t,\tau\rangle_{\mathbb{R}^3})^{3/2}}\right]|_{\rho=\rho_1}\ ,$$

while in case of SGG

$$h_2(t,\tau) := (4\pi R_0)^{-1}\frac{\partial^2}{\partial\rho^2}\left[\frac{\rho^2 - R_0^2}{(\rho^2 + R_0^2 - 2\,\langle t,\tau\rangle_{\mathbb{R}^3})^{3/2}}\right]|_{\rho=\rho_2}\ ,$$

and the surface of the Earth Ω_{R_0} and satellite orbits Ω_{ρ_i} are assumed to be concentric spheres of radii R_0 and ρ_i, $R_0 < \rho_i$.

The joint regularization of multiple observation models (1), (2) can be formulated as an optimization that involves minimization of an objective functional $\Phi(x)$, which combines the measures of data misfit $\|A_i x - y_i^{e_i}\|_{\mathcal{Y}_i}$, $i \in \mathbf{N}_m$, with a regularization measure. In the spirit of the Tikhonov-Phillips regularization the latter can be chosen as the norm $\|\cdot\|_{\mathcal{X}}$ of the solution space. In this way, the objective functional can be written as

$$\Phi(x) := \sum_{i\in\mathbf{N}_m} \lambda_i \|A_i x - y_i^{e_i}\|_{\mathcal{Y}_i}^2 + \|x\|_{\mathcal{X}}^2, \tag{4}$$

where $\lambda_i \in (0,\infty)$ are the regularization parameters. The multiple regularization parameters are introduced in regularization problem (4) to adjust contributions for data misfit from different observation models.

It is convenient to rewrite the objective functional $\Phi(x)$ in (4) in a compact form by introducing a direct sum space of the observation spaces. To this end, for each $i \in \mathbf{N}_m$, we use $\mathcal{Y}_{i,\lambda_i}$ to denote the observation space $\mathcal{Y}_i$ equipped with the scaled inner product $\langle\cdot,\cdot\rangle_{\mathcal{Y}_{i,\lambda_i}} := \lambda_i \langle\cdot,\cdot\rangle_{\mathcal{Y}_i}$, and define the weighted direct sum $\mathbb{Y}_{\boldsymbol{\lambda}}$ of the observation spaces $\mathcal{Y}_{i,\lambda_i}$ by

$$\mathbb{Y}_{\boldsymbol{\lambda}} := \bigoplus_{i\in\mathbf{N}_m} \mathcal{Y}_{i,\lambda_i},$$

where $\boldsymbol{\lambda} := (\lambda_i : i \in \mathbf{N}_m)$, and for $\boldsymbol{u} := (u_i : i \in \mathbf{N}_m) \in \mathbb{Y}_{\boldsymbol{\lambda}}$,

$$\|\boldsymbol{u}\|_{\mathbb{Y}_{\boldsymbol{\lambda}}}^2 = \sum_{i\in\mathbf{N}_m} \lambda_i \|u_i\|_{\mathcal{Y}_i}^2.$$

By letting $\boldsymbol{y}^{\boldsymbol{e}} := (y_i^{e_i} : i \in \mathbf{N}_m) \in \mathbb{Y}_{\boldsymbol{\lambda}}$, and $\mathbb{A}_{\boldsymbol{\lambda}} x := (A_i x : i \in \mathbf{N}_m) \in \mathbb{Y}_{\boldsymbol{\lambda}}$, the objective functional (4) may be represented as

$$\Phi(x) = \|\mathbb{A}_{\boldsymbol{\lambda}} x - \boldsymbol{y}^{\boldsymbol{e}}\|_{\mathbb{Y}_{\boldsymbol{\lambda}}}^2 + \|x\|_{\mathcal{X}}^2. \tag{5}$$

Note that $\mathbb{A}_{\boldsymbol{\lambda}}$ is a compact linear injective operator from $\mathcal{X}$ to $\mathbb{Y}_{\boldsymbol{\lambda}}$.

In the representation (5), we can conveniently obtain the classical Tikhonov-Phillips form of the minimizer $x_{\boldsymbol{\lambda}}^{\boldsymbol{e}}$ of Φ defined by (4) in terms of $\mathbb{A}_{\boldsymbol{\lambda}}$. Specifically, the minimizer $x_{\boldsymbol{\lambda}}^{\boldsymbol{e}}$ of Φ satisfies the equation

$$x_{\boldsymbol{\lambda}}^{\boldsymbol{e}} = (I + \mathbb{A}_{\boldsymbol{\lambda}}^* \mathbb{A}_{\boldsymbol{\lambda}})^{-1} \mathbb{A}_{\boldsymbol{\lambda}}^* \boldsymbol{y}^{\boldsymbol{e}},$$

where $I : \mathcal{X} \to \mathcal{X}$ is the identity operator and $\mathbb{A}_{\boldsymbol{\lambda}}^* : \mathbb{Y}_{\boldsymbol{\lambda}} \to \mathcal{X}$ is the adjoint of $\mathbb{A}_{\boldsymbol{\lambda}}$. Clearly, for any $\boldsymbol{u}$ the adjoint operator $\mathbb{A}_{\boldsymbol{\lambda}}^*$ has the form

$$\mathbb{A}_{\boldsymbol{\lambda}}^* \boldsymbol{u} = \sum_{i\in\mathbf{N}_m} \lambda_i A_i^* u_i.$$

From the representation (5) it follows that the regularized approximant $x = x_{\boldsymbol{\lambda}}^{\boldsymbol{e}}$ is the solution of the equation

$$x + \sum_{i\in\mathbf{N}_m} \lambda_i A_i^* A_i x = \sum_{i\in\mathbf{N}_m} \lambda_i A_i^* y_i^{e_i}. \tag{6}$$

Equation (6) was obtained in [17] by the Bayesian reasoning. It is also suggested in [17] to relate the values of the regularization parameters λ_i with the observation noise levels (variances) ε_i as follows:

$$\lambda_i = \lambda_1 \frac{\varepsilon_1^2}{\varepsilon_i^2},\ i \in \mathbf{N}_m. \tag{7}$$

Note that for the given noise levels this relation reduces the multi-parameter regularization (4) to a single parameter one, since only λ_1 needs to be chosen.

The heuristic rule (7) can be derived from a bound for the noise propagation error. Specifically, we have that

$$\begin{aligned}\|x^0_{\boldsymbol{\lambda}} - x^{\boldsymbol{e}}_{\boldsymbol{\lambda}}\|_{\mathcal{X}} &= \|(I + \mathbb{A}^*_{\boldsymbol{\lambda}}\mathbb{A}_{\boldsymbol{\lambda}})^{-1}\mathbb{A}^*_{\boldsymbol{\lambda}}(\boldsymbol{y^0} - \boldsymbol{y^e})\|_{\mathcal{X}} \\ &\leqslant \|(I + \mathbb{A}^*_{\boldsymbol{\lambda}}\mathbb{A}_{\boldsymbol{\lambda}})^{-1}\mathbb{A}^*_{\boldsymbol{\lambda}}\|_{\mathbb{Y}_{\boldsymbol{\lambda}}\to\mathcal{X}}\|\boldsymbol{y^0} - \boldsymbol{y^e}\|_{\mathbb{Y}_{\boldsymbol{\lambda}}} \\ &\leqslant \frac{1}{2}\|\mathbb{A}_{\boldsymbol{\lambda}}x^\dagger - \boldsymbol{y^e}\|_{\mathbb{Y}_{\boldsymbol{\lambda}}}\end{aligned}$$

where $\boldsymbol{y^0} := (y^0_i : i \in \mathbf{N}_m) = (A_i x^\dagger : i \in \mathbf{N}_m) \in \mathbb{Y}_{\boldsymbol{\lambda}}$. It follows from Assumption (2) that

$$\|x^0_{\boldsymbol{\lambda}} - x^{\boldsymbol{e}}_{\boldsymbol{\lambda}}\|_{\mathcal{X}} \leqslant \frac{1}{2}\left(\sum_{i\in\mathbf{N}_m}\lambda_i\varepsilon_i^2\right)^{\frac{1}{2}}. \tag{8}$$

The heuristics behind the rule (7) is clear: The rule equates all the terms from the bound (8) and balances data misfits against each other. Then the final balance may be achieved by making a choice of the last remaining parameter $\lambda = \lambda_1$. The latter can be chosen by known single-parameter choice rules such as the quasi-optimality criterion [28]. We label the above heuristics as "M1" and provide an algorithm (Algorithm 2.1) in pseudo-code to facilitate the usage.

Algorithm 2.1 for M1

Input: $A_i,\ y_i^{e_i}, \varepsilon_i,\ i \in \mathbf{N}_m,\ \Sigma$, where Σ is the parameter set
Output: $\tilde{x}$

for $j = 1$ to $|\Sigma|$, where $\lambda_1^{(j)} \in \Sigma$ **do**
  for $i = 1$ to m **do**
    $\lambda_i^{(j)} \leftarrow \lambda_1^{(j)}\varepsilon_1/\varepsilon_i$
  end for
  $x^{(j)} \leftarrow (I + \mathbb{A}^*_{\boldsymbol{\lambda}^{(j)}}\mathbb{A}_{\boldsymbol{\lambda}^{(j)}})^{-1}\mathbb{A}^*_{\boldsymbol{\lambda}^{(j)}}\boldsymbol{y^e}$, where $\boldsymbol{\lambda}^{(j)} := (\lambda_i^{(j)} : i \in \mathbf{N}_m)$
  if $j \geqslant 2$ **then**
    $\Delta^{(j)} \leftarrow \|x^{(j)} - x^{(j-1)}\|_{\mathcal{X}}$
  end if
end for
$j_* \leftarrow \arg\min\{\Delta^{(j)},\ 2 \leqslant j \leqslant |\Sigma|\}$
return $\tilde{x} \leftarrow x^{(j_*)}$

There are alternatives to (7), for example, a multiple version of the well-known quasi-optimality criterion [28]. For the sake of the clarity of the presentation we describe it here only for the case of two observation equations (1) ($m = 2$), which we label as "M2". Algorithm 2.2 is provided to illustrate the mechanism behind it. There are also several other ways of selecting the values of the regularization parameters in (4) and (6). In the next section we shall discuss how we can gain from a variety of rules.

Algorithm 2.2 for M2

Input: $A_i,\ y_i^{e_i},\ i = 1, 2,\ \Sigma$
Output: $\tilde{x}$

for $j = 1$ to $|\Sigma|$, where $\lambda_1^{(j)} \in \Sigma$ **do**
  for $k = 1$ to $|\Sigma|$, where $\lambda_2^{(k)} \in \Sigma$ **do**
    $x^{(j,k)} \leftarrow (I + \mathbb{A}^*_{\boldsymbol{\lambda}^{(j,k)}}\mathbb{A}_{\boldsymbol{\lambda}^{(j,k)}})^{-1}\mathbb{A}^*_{\boldsymbol{\lambda}^{(j,k)}}\boldsymbol{y^e}$, where $\boldsymbol{\lambda}^{(j,k)} = (\lambda_1^{(j)}, \lambda_2^{(k)})$
    if $k \geqslant 2$ **then**
      $\Delta^{(j,k)} \leftarrow \|x^{(j,k)} - x^{(j,k-1)}\|_{\mathcal{X}}$
    end if
  end for
end for
$(j_*, k_*) \leftarrow \arg\min\{\Delta^{(j,k)},\ 1 \leqslant j \leqslant |\Sigma|,\ 2 \leqslant k \leqslant |\Sigma|\}$
return $\tilde{x} \leftarrow x^{(j_*,k_*)}$

3 Aggregation By a Linear Functional Strategy

A critical issue in solving the joint inversion problem is the choice of the multiple regularization parameters involved in the model. We may propose various rules for the choice of the weighted parameters of multiple observation spaces if certain a prior solution information is available. In the case of less of dominant criteria, a feasible way to solve the joint inversion problem is to make use of the variety of parameters, or the resulting solution candidates. In this sense choosing certain linear combination of multiple solutions, which result from different parameter choices, as a new solution to the joint inversion problem may be more efficient. This section is devoted to developing such a method. We propose an "aggregation" method in a more general sense with the joint inversion problem as a special example. Specifically, we describe the construction of an aggregator and discuss technical difficulties in its realization. We then propose a linear functional strategy to resolve the problem and estimate the related errors. Furthermore, we show that the parameter choice strategy by using the balancing principle can achieve almost the best guaranteed accuracy.

We begin with the review of the aggregation. We assume that there are n various rules available for choosing the multiple parameters $\boldsymbol{\lambda}$, which result in n different approximations $\tilde{x}_j \in \mathcal{X}$, $j \in \mathbf{N}_n$, to the element of interest $x^\dagger$. In practice, usually we do not know which of these approximations better fits the element $x^\dagger$. Moreover, the approximations $\tilde{x}_j$, $j \in \mathbf{N}_n$, may complement each other. An attractive way to resolve the arising uncertainty is to "aggregate" these approximations. That is, we find the best linear combination

$$x_* := \sum_{j \in \mathbf{N}_n} \beta_j^* \tilde{x}_j,$$

in the sense that x_* solves the minimization problem

$$\|x^\dagger - x_*\|_{\mathcal{X}} = \min_{\beta_j} \left\| x^\dagger - \sum_{j \in \mathbf{N}_n} \beta_j \tilde{x}_j \right\|_{\mathcal{X}}. \tag{9}$$

The solution x_* of (9) is determined by $\boldsymbol{\beta}_* := (\beta_j^* : j \in \mathbf{N}_n)^T$, which is the solution of a system of linear equations. To observe this, we need the inner product $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ of Hilbert space $\mathcal{X}$. Letting

$$\boldsymbol{G} := (\langle \tilde{x}_r, \tilde{x}_s \rangle_{\mathcal{X}} : r, s \in \mathbf{N}_n)$$

and introducing $\boldsymbol{\beta} := (\beta_j : j \in \mathbf{N}_n)^T$ and $\boldsymbol{\kappa} := (\kappa_j : j \in \mathbf{N}_n)^T$, where $\kappa_j := \left\langle \tilde{x}_j, x^\dagger \right\rangle_{\mathcal{X}}$, $j \in \mathbf{N}_n$, we see that $\boldsymbol{\beta}_*$ solves the linear system

$$\boldsymbol{G}\boldsymbol{\beta} = \boldsymbol{\kappa}. \tag{10}$$

If we assume, without loss of generality, that approximations $\tilde{x}_j$, $j \in \mathbf{N}_n$, are linearly independent, then the Gram matrix $\boldsymbol{G}$ is positive definite and thus invertible such that

$$\|\boldsymbol{G}^{-1}\|_{\mathbb{R}^n \to \mathbb{R}^n} \leqslant \gamma, \tag{11}$$

where γ is a constant that depends only on $\tilde{x}_j$, $j \in \mathbf{N}_n$. Thus, $\boldsymbol{\beta}_*$ has the representation

$$\boldsymbol{\beta}_* = \boldsymbol{G}^{-1}\boldsymbol{\kappa}$$

and

$$x_* = \tilde{\boldsymbol{x}}\boldsymbol{\beta}_* := \sum_{j \in \mathbf{N}_n} \beta_j^* \tilde{x}_j,$$

where $\tilde{\boldsymbol{x}} := (\tilde{x}_j : j \in \mathbf{N}_n)$.

System (10) cannot be solved without knowing the vector $\boldsymbol{\kappa}$. However, the vector $\boldsymbol{\kappa}$ involves the unknown solution $x^\dagger$. Therefore, we turn to finding an approximation of system (10). A possible approach is to approximate $\boldsymbol{\kappa}$ by a $\tilde{\boldsymbol{\kappa}} := (\tilde{\kappa}_j : j \in \mathbf{N}_n)^T$, and to get

$$\tilde{\boldsymbol{\beta}} = \boldsymbol{G}^{-1}\tilde{\boldsymbol{\kappa}}$$

by solving the system

$$\boldsymbol{G}\boldsymbol{\beta} = \tilde{\boldsymbol{\kappa}}. \tag{12}$$

This yields an effective aggregator

$$x_{ag} := \tilde{\boldsymbol{x}}\tilde{\boldsymbol{\beta}} = \sum_{j \in \mathbf{N}_n} \tilde{\beta}_j \tilde{x}_j, \tag{13}$$

of $\tilde{x}_j$, $j \in \mathbf{N}_n$. In other words, we find an approximation $\tilde{\boldsymbol{\beta}} := (\tilde{\beta}_j : j \in \mathbf{N}_n)^T$ of $\boldsymbol{\beta}_*$ such that $x_{ag} = \tilde{\boldsymbol{x}}\tilde{\boldsymbol{\beta}}$ is "nearly as good as" x_*. Following this idea, we describe a construction of $\tilde{\boldsymbol{\kappa}}$ and estimate the resulting errors.

3.1 Selected results on the linear functional strategy

In this subsection we present an approach to approximate $\tilde{\boldsymbol{\kappa}}$ by using the linear functional strategy as introduced in [1, 2, 19]. The advantage of this strategy is that if one is not interested in completely knowing $x^\dagger$, but instead in only some quantity derived from it, such as the value of a bounded linear functional $\tilde{x}(\cdot) = \langle \tilde{x}, \cdot \rangle_{\mathcal{X}}$ of the solution $x^\dagger$, then this quantity can be estimated more accurately than the solution $x^\dagger$ in $\mathcal{X}$.

It is reasonable to measure the closeness between an effective aggregator x_{ag} and the ideal approximant x_* in terms of the accuracy of the best reconstruction of $x^\dagger$ from the most trustable observation equation

$$y^e = Ax + e. \tag{14}$$

Model (14) is selected from the set of the considered models (1) according to one of the following rules: (a) The operator $A \in \{A_i : i \in \mathbf{N}_m\}$ makes the corresponding inverse problem less ill-posed than the other design operators, or (b) the data $y^e \in \{y_i^{e_i} : i \in \mathbf{N}_m\}$ are provided with the smallest noise level $\varepsilon := \min\{\varepsilon_i : i \in \mathbf{N}_m\}$ such that

$$\|Ax^\dagger - y^e\|_{\mathcal{Y}} \leqslant \varepsilon, \tag{15}$$

where $\mathcal{Y}$ is the corresponding observation space $\mathcal{Y}_i$, for some $i \in \mathbf{N}_m$. Then the best guaranteed accuracy of the reconstruction of $x^\dagger \in \mathcal{X}$ from (14) and (15) can be expressed in terms of the noise level ε and the smoothness of $x^\dagger$. From [24], the smoothness of $x^\dagger$ can be represented in the form of the source condition to be described below. According to [24], a function $\varphi : [0, \|A\|^2_{\mathcal{X}\to\mathcal{Y}}] \to [0, \infty)$ is called an index function if it is continuous, strictly increasing, and satisfies $\varphi(0) = 0$. Suppose that an index function φ is given. Let σ_k, w_k, $k \in \mathbb{N}$ be the singular values and the corresponding singular vectors of A^*A, such that w_k, $k \in \mathbb{N}$, form a standard orthogonal system in $\mathcal{X}$. We assume that the following source condition holds

$$x^\dagger \in A_\varphi(R) := \left\{ x \in \mathcal{X},\ x = \varphi(A^*A)v := \sum_{k\in\mathbb{N}} \varphi(\sigma_k) \langle v, w_k \rangle_{\mathcal{X}} w_k,\ \|v\|_{\mathcal{X}} \leqslant R \right\}, \tag{16}$$

where R is a positive number. It is also known from [13, 25] that

$$\inf_{L:\mathcal{Y}\to\mathcal{X}} \sup_{x^\dagger \in A_\varphi(R)} \sup_{\substack{y^e: \\ \|Ax^\dagger - y^e\|_{\mathcal{Y}} \leqslant \varepsilon}} \|x^\dagger - Ly^e\|_{\mathcal{X}} = O(\varphi(\theta^{-1}(\varepsilon))), \tag{17}$$

where $\theta(t) := \varphi(t)\sqrt{t}$ for $t \geqslant 0$, and the infimum is taken over all possible mappings L from $\mathcal{Y}$ to $\mathcal{X}$. Clearly, formula (17) ensures that the best guaranteed accuracy of the reconstruction of $x^\dagger \in A_\varphi(R)$ from the observation (14) and (15) has the order of $\varphi(\theta^{-1}(\varepsilon))$.

For the purpose of analyzing the Tikhonov-Phillips regularization and its multi-parameter version, such as (4), it is natural to assume that the source condition (16) is generated by a given index function φ that is covered by the qualification, denoted by p, of the Tikhonov-Phillips method (i.e., $p = 1$) in the sense of [25]. That is to require that the function $\omega(t) := \frac{t^p}{\varphi(t)}$, where $p = 1$ in case of Tikhonov-Phillips regularization, is a nondecreasing function. Then from [25] it is known that the best guaranteed order of accuracy can be achieved within the Tikhonov-Phillips scheme. Recall that for a regularization parameter α, the Tikhonov-Phillips regularization of (14) and (15) is defined as

$$x_\alpha = \arg\min\{\|Ax - y^e\|^2_{\mathcal{Y}} + \alpha\|x\|_{\mathcal{X}}\} = (\alpha I + A^*A)^{-1}A^*y^e.$$

We recall below a result from [25].

Theorem 3.1. *If φ is an index function such that the function $\omega(t) = \frac{t}{\varphi(t)}$ is nondecreasing, then for $\alpha = \theta^{-1}(\varepsilon)$*

$$\sup_{x^\dagger \in A_\varphi(R)} \sup_{\substack{y^e: \\ \|Ax^\dagger - y^e\|_{\mathcal{Y}} \leqslant \varepsilon}} \|x^\dagger - x_\alpha\|_{\mathcal{X}} = O(\varphi(\theta^{-1}(\varepsilon))).$$

Since the Hilbert space $\mathcal{X}$ is self-adjoint, an element $\tilde{x}$ representing a bounded linear functional $\tilde{x}(x) := \langle \tilde{x}, x\rangle_{\mathcal{X}}$ has certain order of smoothness that can be expressed in terms of the source condition $\tilde{x} \in Range(\psi(A^*A))$ with an index function ψ. If we use the Tikhonov-Phillips regularization to approximate the value $\tilde{x}(x^\dagger) := \langle \tilde{x}, x^\dagger\rangle_{\mathcal{X}}$ by $\langle \tilde{x}, x_\alpha\rangle_{\mathcal{X}}$, then the following result is known (see, e.g., [5], [20], Remark 2.6).

Theorem 3.2. *If $\tilde{x} \in Range(\psi(A^*A))$ and the index functions φ and ψ are such that the functions $\frac{\sqrt{t}}{\psi(t)}$ and $\frac{t}{\varphi(t)\psi(t)}$ are nondecreasing, then for $\alpha = \theta^{-1}(\varepsilon)$*

$$\sup_{x^\dagger \in A_\varphi(R)} \sup_{\substack{y^\varepsilon: \\ \|Ax^\dagger - y^\varepsilon\|_{\mathcal{Y}} \leqslant \varepsilon}} \left| \left\langle \tilde{x}, x^\dagger \right\rangle_{\mathcal{X}} - \langle \tilde{x}, x_\alpha\rangle_{\mathcal{X}} \right| = O(\varphi(\theta^{-1}(\varepsilon))\psi(\theta^{-1}(\varepsilon))) = o(\varphi(\theta^{-1}(\varepsilon))).$$

Comparing the above theorems, we conclude that potentially the value of $\langle \tilde{x}, x^\dagger\rangle_{\mathcal{X}}$ allows a more accurate estimation than the solution $x^\dagger$, provided that $\tilde{x}$ satisfies the hypothesis of Theorem 3.2. In the next theorem, we show that for the purpose of aggregation the hypothesis of Theorem 3.2 is not too restrictive.

Theorem 3.3. *If φ is an index function such that the function $\omega(t) = \frac{t}{\varphi(t)}$ is increasing, $\omega(0) := \lim_{t\to 0^+} \omega(t) = 0$ and assume that $x^\dagger \in A_\varphi(R)$, then for any $\tilde{x} \in \mathcal{X}$ there is an index function ψ such that $\tilde{x} \in Range(\psi(A^*A))$ and the functions $\frac{\sqrt{t}}{\psi(t)}$ and $\frac{t}{\varphi(t)\psi(t)}$ are nondecreasing.*

Proof: Consider the compact linear operator $B := \omega(A^*A)$ which is self-adjoint, injective and non-negative. From Corollary 2 of [24], it follows that there is a concave index function ψ_0 and a constant $R' > 0$ such that $\tilde{x} = \psi_0(B)v_0$, $\|v_0\|_{\mathcal{X}} \leqslant R'$. Thus, $\tilde{x}$ admits a representation

$$\tilde{x} = (\psi_0(B))^{\frac{1}{2}}(\psi_0(B))^{\frac{1}{2}}v_0 = \psi_0^{\frac{1}{2}}(\omega(A^*A))v, \tag{18}$$

where $v := (\psi_0(B))^{\frac{1}{2}}v_0 \in \mathcal{X}$.

Next, we consider the function $\psi(t) := (\psi_0(\omega(t)))^{\frac{1}{2}}$ and prove that it has the desired properties described in this theorem. It follows from (18) that $\tilde{x} \in Range(\psi(A^*A))$. Moreover, in view of the concavity of the index function ψ_0, we observe that for any $0 < u_1 < u_2$

$$\frac{u_1}{\psi_0(u_1)} \leqslant \frac{u_2}{\psi_0(u_2)}. \tag{19}$$

Hence, for $0 < t_1 < t_2$ we have that $\omega(t_1) < \omega(t_2)$ and

$$\begin{aligned}\frac{t_1}{\varphi(t_1)\psi(t_1)} &= \left(\frac{t_1}{\varphi(t_1)}\right)^{\frac{1}{2}}\left(\frac{t_1}{\varphi(t_1)\psi_0(\omega(t_1))}\right)^{\frac{1}{2}} = (\omega(t_1))^{\frac{1}{2}}\left(\frac{\omega(t_1)}{\psi_0(\omega(t_1))}\right)^{\frac{1}{2}} \\ &\leqslant (\omega(t_2))^{\frac{1}{2}}\left(\frac{\omega(t_2)}{\psi_0(\omega(t_2))}\right)^{\frac{1}{2}} = \frac{t_2}{\varphi(t_2)\psi(t_2)},\end{aligned}$$

where we have used (19) with $u_1 := \omega(t_1)$, $u_2 := \omega(t_2)$ and the increasing monotonicity of ω. This proves that the function $\frac{t}{\varphi(t)\psi(t)}$ is nondecreasing. The same feature of the function $\frac{t^{\frac{1}{2}}}{\psi(t)}$ may be proved in the same way. Indeed, for any $0 < t_1 < t_2$ we have that

$$\begin{aligned}\frac{t_1^{\frac{1}{2}}}{\psi(t_1)} &= \left(\frac{t_1}{\psi_0(\omega(t_1))}\right)^{\frac{1}{2}} = \left(\frac{\omega(t_1)}{\psi_0(\omega(t_1))}\right)^{\frac{1}{2}}\varphi^{\frac{1}{2}}(t_1) \\ &\leqslant \left(\frac{\omega(t_2)}{\psi_0(\omega(t_2))}\right)^{\frac{1}{2}}\varphi^{\frac{1}{2}}(t_2) = \frac{t_2^{\frac{1}{2}}}{\psi^2(t_2)},\end{aligned}$$

proving the desired result. □

3.2 Application in aggregation

In the framework of the linear functional strategy described earlier, for each $j \in \mathbf{N}_n$, we can approximate the component κ_j of the vector $\boldsymbol{\kappa}$ by

$$\tilde{\kappa}_j = \left\langle \tilde{x}_j, x_{\alpha_j}\right\rangle_{\mathcal{X}}, \tag{20}$$

where $x_{\alpha_j} := (\alpha_j I + A^*A)^{-1}A^*y^e,\ j \in \mathbf{N}_n$ are feasible substitutes for $x^\dagger$. Combining Theorems 3.2 and 3.3, we see that for $x^\dagger \in Range(\varphi(A^*A))$ there exists an index function ψ_j such that $\tilde{x}_j \in Range(\psi_j(A^*A))$ and

$$\left|\left\langle \tilde{x}_j, x^\dagger \right\rangle_{\mathcal{X}} - \langle \tilde{x}_j, x_{\alpha_j} \rangle_{\mathcal{X}}\right| = O(\varphi(\theta^{-1}(\varepsilon))\psi_j(\theta^{-1}(\varepsilon))) = o(\varphi(\theta^{-1}(\varepsilon))), \tag{21}$$

where $\alpha_j = \alpha = \theta^{-1}(\varepsilon)$. Theoretically, the order of accuracy $o(\varphi(\theta^{-1}(\varepsilon)))$ in approximating κ_j can be achieved by applying the linear functional strategy with the same value of the regularization parameter α. However, in practice, the function φ describing the smoothness of the unknown solution $x^\dagger$ is unknown. As a result, one cannot implement *a priori* parameter choice $\alpha_j = \alpha = \theta^{-1}(\varepsilon)$. In principle, this difficulty may be resolved by use of the so-called Lepskii-type balancing principle, introduced in [5, 11] in the contest of the linear functional strategy. But, *a posteriori* parameter choice strategy presented in those papers requires the knowledge of the index functions ψ_j describing the smoothness of $\tilde{x}_j$ in terms of the source condition $\tilde{x}_j \in Range(\psi_j(A^*A))$. This requirement may be restrictive in some applications. The proof of Theorem 3.3, for example, gives the formula $\psi_j(t) = (\psi_0(\omega(t)))^{\frac{1}{2}}$, where both functions ψ_0, ω depend on the unknown index function φ. To overcome this difficulty, we consider below a modification of the balancing principle to achieve the error bounds (21) without requiring the knowledge of φ and ψ_j.

The theory of the balancing principle is known in the literature (see, for example, [11], [20] Section 1.1.5, and [23]). Following the general theory, we formulate a version of the balancing principle suitable for our context. In view of the representation

$$\begin{aligned}\left\langle \tilde{x}_j, x^\dagger \right\rangle_{\mathcal{X}} - \langle \tilde{x}_j, x_\alpha \rangle_{\mathcal{X}} &= \left\langle \tilde{x}_j, x^\dagger \right\rangle_{\mathcal{X}} - \langle \tilde{x}_j, (\alpha I + A^*A)^{-1}A^*y^e \rangle_{\mathcal{X}} \\ &= \alpha \left\langle (\alpha I + A^*A)^{-1}\tilde{x}_j, x^\dagger \right\rangle_{\mathcal{X}} - \langle e, (\alpha I + AA^*)^{-1}A\tilde{x}_j \rangle_{\mathcal{Y}}, \end{aligned} \tag{22}$$

which is well known as the decomposition of the error into the noise-free term and the noise term, we consider the functions of α

$$\begin{aligned} E_j(\alpha) &:= \left|\left\langle \tilde{x}_j, x^\dagger \right\rangle_{\mathcal{X}} - \langle \tilde{x}_j, x_\alpha \rangle_{\mathcal{X}}\right|, \\ B_j(\alpha) &:= \alpha \left|\left\langle (\alpha I + A^*A)^{-1}\tilde{x}_j, x^\dagger \right\rangle_{\mathcal{X}}\right| \end{aligned} \tag{23}$$

and

$$V_j(\alpha) := \left\|(\alpha I + AA^*)^{-1}A\tilde{x}_j\right\|_{\mathcal{Y}}. \tag{24}$$

From (15), (22) we can derive the following bound in terms of $V_j(\alpha), B_j(\alpha)$:

$$E_j(\alpha) \leqslant B_j(\alpha) + \varepsilon V_j(\alpha). \tag{25}$$

At the same time, it should be noted that $V_j(\alpha)$ is a monotonically decreasing continuous function of α. Following the analysis given above, we propose a modified version of the balancing principle for the data functional strategy on Tikhonov-Phillips regularization.

Consider the Tikhonov-Phillips regularized data functional $\langle \tilde{x}_j, x_\alpha \rangle_{\mathcal{X}}$. We choose the value of α from the finite set

$$\Sigma_N := \{\varepsilon^2 = \alpha^{(1)} < \alpha^{(2)} < \ldots < \alpha^{(N)} = 1\}$$

according to the balancing principle

$$\alpha = \alpha_j := \max\{\alpha^{(k)} \in \Sigma_N : \left|\langle \tilde{x}_j, x_{\alpha^{(k)}} \rangle_{\mathcal{X}} - \langle \tilde{x}_j, x_{\alpha^{(l)}} \rangle_{\mathcal{X}}\right| \leqslant 4\varepsilon V_j(\alpha^{(l)}),\ l \in \mathbf{N}_k\}. \tag{26}$$

Theorem 3.4. *If the value of α is chosen from the finite set Σ_N according to the balancing principle (26), then*

$$E_j(\alpha_j) \leqslant C \min_{\alpha \in [\varepsilon^2, 1]} \{B_j(\alpha) + \varepsilon V_j(\alpha)\}, \tag{27}$$

where the coefficient C can be estimated as

$$C := 6 \max\{V_j(\alpha^{(l+1)})/V_j(\alpha^{(l)}),\ l \in \mathbf{N}_{N-1}\}.$$

Theorem 3.4 can be proved by the same argument as in [20] (Section 1.1.5) and [23]. It leads to the following result.

Theorem 3.5. *Suppose that $x^\dagger \in A_\varphi(R)$ and φ is such that the function $\omega(t) = \frac{t}{\varphi(t)}$ is increasing and $\omega(0) = 0$. If $\alpha = \alpha_j$ is chosen according to (26) and the bound (15) holds true then*

$$E_j(\alpha_j) = o(\varphi(\theta^{-1}(\varepsilon))).$$

Proof: We shall use (25) to estimate $E_j(\alpha_j)$. We first consider (24). By the assumption of this theorem on the index function φ, using the proof of Proposition 2.15 of [20], we observe that for $x^\dagger \in A_\varphi(R)$, the assumption of Theorem 3.3 is satisfied and we have that

$$V_j(\alpha) \leqslant C\frac{\psi_j(\alpha)}{\sqrt{\alpha}}, \tag{28}$$

where C depends only on the norm of $\tilde{x}_j$ and ψ_j is an index function satisfying the assumption of Theorem 3.2 in a sense that $\tilde{x}_j \in Range(\psi_j(A^*A))$, and the both functions $\frac{\sqrt{t}}{\psi_j(t)}$ and $\frac{t}{\varphi(t)\psi_j(t)}$ are nondecreasing ones. The existence of such ψ_j is guaranteed by Theorem 3.3.

We then estimate (23). Again, from the proof of Proposition 2.15 of [20], the following bound

$$B_j(\alpha) \leqslant C\varphi(\alpha)\psi_j(\alpha) \tag{29}$$

is satisfied, where the coefficient C depends only on the norms $x^\dagger, \tilde{x}_j$. Theorem 3.4 tells us that if we choose the value of α as per (26) then the error estimate (27) can be achieved. Thus, in view of the bounds (27), (29) and (28) it is clear that $\theta^{-1}(\varepsilon) \in [\varepsilon^2, 1]$ and

$$\begin{aligned}\min_{\alpha\in[\varepsilon^2,1]}\{B_j(\alpha)+\varepsilon V_j(\alpha)\} &\leqslant C\min_{\alpha\in[\varepsilon^2,1]}\left\{\varphi(\alpha)\psi_j(\alpha)+\varepsilon\frac{\psi_j(\alpha)}{\sqrt{\alpha}}\right\}\\ \leqslant C&\left(\varphi(\theta^{-1}(\varepsilon))\psi_j(\theta^{-1}(\varepsilon))+\varepsilon\frac{\psi_j(\theta^{-1}(\varepsilon))}{\sqrt{\theta^{-1}(\varepsilon)}}\right) = 2C\varphi(\theta^{-1}(\varepsilon))\psi_j(\theta^{-1}(\varepsilon)),\end{aligned} \tag{30}$$

proving the desired result. □

It is worth to note that given A and $\tilde{x}_j$ the value of $V_j(\alpha)$ at any point $\alpha \in (0, 1]$ can be calculated directly as the norm of the solution of the equation $\alpha u + AA^*u = A\tilde{x}_j$ in the space $\mathcal{Y}$, and it does not require any knowledge of ψ_j.

We prove the following main result of aggregation.

Theorem 3.6. *Suppose that x_* is the ideal aggregator in the sense of (9) and x_{ag} is its approximant constructed according to (13), (12) and (20). Let y^e be an observation that satisfies the bound (15). If $x^\dagger \in A_\varphi(R)$, where φ is an index function such that the function $\omega(t) = \frac{t}{\varphi(t)}$ is increasing, $\omega(0) = 0$, and the values of the regularization parameters α_j, $j \in \mathbf{N}_n$ in (20) are chosen according to the balancing principle (26) using only y^e, A and $\tilde{x}_j$, $j \in \mathbf{N}_n$, then*

$$\|x^\dagger - x_{ag}\|_{\mathcal{X}} - \|x^\dagger - x_*\|_{\mathcal{X}} = o(\varphi(\theta^{-1}(\varepsilon))).$$

Proof: According to the triangular inequality, we have that

$$\|x^\dagger - x_{ag}\|_{\mathcal{X}} - \|x^\dagger - x_*\|_{\mathcal{X}} \leqslant \|x_* - x_{ag}\|_{\mathcal{X}} = \|\tilde{\boldsymbol{x}}(\boldsymbol{\beta}_* - \tilde{\boldsymbol{\beta}})\|_{\mathcal{X}} \leqslant \|\tilde{\boldsymbol{x}}\|_{\mathcal{X}^n}\|\boldsymbol{\beta}_* - \tilde{\boldsymbol{\beta}}\|_{\mathbb{R}^n}.$$

It remains to estimate $\|\boldsymbol{\beta}_* - \tilde{\boldsymbol{\beta}}\|_{\mathbb{R}^n}$. To this end, from (27), (30) we observe that

$$\|\boldsymbol{\kappa} - \tilde{\boldsymbol{\kappa}}\|_{\mathbb{R}^n} = \left(\sum_{j\in\mathbf{N}_n}\left|\left\langle\tilde{x}_j, x^\dagger\right\rangle_{\mathcal{X}} - \left\langle\tilde{x}_j, x_{\alpha_j}\right\rangle_{\mathcal{X}}\right|^2\right)^{\frac{1}{2}} \leqslant C\sqrt{n}\varphi(\theta^{-1}(\varepsilon))\max_j\psi_j(\theta^{-1}(\varepsilon)),$$

where C depends only on the norms of $x^\dagger$ and $\tilde{x}_j$, $j \in \mathbf{N}_n$. Thus, we get that

$$\|\boldsymbol{\kappa} - \tilde{\boldsymbol{\kappa}}\|_{\mathbb{R}^n} = o(\varphi(\theta^{-1}(\varepsilon)).$$

It follows from (11) that

$$\|\boldsymbol{\beta}_* - \tilde{\boldsymbol{\beta}}\|_{\mathbb{R}^n} = \|\boldsymbol{G}^{-1}(\boldsymbol{\kappa} - \tilde{\boldsymbol{\kappa}})\|_{\mathbb{R}^n} \leqslant \gamma\|\boldsymbol{\kappa} - \tilde{\boldsymbol{\kappa}}\|_{\mathbb{R}^n} = o(\varphi(\theta^{-1}(\varepsilon)).$$

Finally, combining the estimates above, we find that

$$\|x^\dagger - x_{ag}\|_{\mathcal{X}} - \|x^\dagger - x_*\|_{\mathcal{X}} \leqslant \|\tilde{\boldsymbol{x}}\|_{\mathcal{X}^n}\|\boldsymbol{\beta}_* - \tilde{\boldsymbol{\beta}}\|_{\mathbb{R}^n} = o(\varphi(\theta^{-1}(\varepsilon))),$$

proving the desired result. □

Theorem 3.6 tells us that the coefficients $\tilde{\beta}_j$ of the aggregator x_{ag} can be effectively obtained from the input data in such a way that the error $\|x^\dagger - x_{ag}\|_{\mathcal{X}}$ differs from the ideal error $\|x^\dagger - x_*\|_{\mathcal{X}}$ by a quantity of higher order than the best guaranteed accuracy of the reconstruction of $x^\dagger$ from the most trustable observation (15).

Note that in our analysis the balancing principle (26) has been used mainly for the theoretical reason. In numerical experiments below the vector $\tilde{\boldsymbol{\beta}}$ of the coefficients of the aggregator x_{ag} is found from the system (12) with (20), where the regularization parameters α_j are chosen by a version of the quasi-optimality criterion which is described below: For the functional $\langle \tilde{x}_j, x^\dagger \rangle_{\mathcal{X}}$, we choose the value $\alpha = \alpha_j$ from

$$\Sigma_N^q = \{\alpha^{(k)} = \alpha^{(0)} q^k : k \in \mathbf{N}_N^0 := 0, 1, \ldots, N\},$$

for some $q > 1$ such that

$$\left| \langle \tilde{x}_j, x_{\alpha_j} \rangle_{\mathcal{X}} - \left\langle \tilde{x}_j, x_{\alpha_j q^{-1}} \right\rangle_{\mathcal{X}} \right| = \min\{ \left| \langle \tilde{x}_j, x_{\alpha^{(k)}} \rangle_{\mathcal{X}} - \langle \tilde{x}_j, x_{\alpha^{(k-1)}} \rangle_{\mathcal{X}} \right| : k \in \mathbf{N}_N \}. \tag{31}$$

Observe that the balancing principle (26) and the version of the quasi-optimality criterion presented above are similar in the sense that both parameter choice rules operate with the differences of the approximate values $\langle \tilde{x}_j, x_\alpha \rangle_{\mathcal{X}}$ of the quantity of interest $\langle \tilde{x}_j, x^\dagger \rangle_{\mathcal{X}}$. In the next section we demonstrate the performance of the aggregation by the linear functional strategy, which is based on the parameter choice rule (31).

To close this subsection, we present the method of aggregation, labelled as "M3", as the third method for the parameter choice. For a series of regularized solutions corresponding to different parameter choices, we do aggregation by using a single observation (15) for the approximation of the objective aggregator with the regularization parameter being chosen by the quasi-optimality method. Algorithm 3.1 is presented below to describe the method.

Algorithm 3.1 for M3

Input: $A, y^e, \tilde{\boldsymbol{x}} := (\tilde{x}_j : \ j \in \mathbf{N}_n), \Sigma$

Output: x_{ag}

for $j = 1$ to n **do**
  for $k = j$ to n **do**
    $G_{k,j} = G_{j,k} \leftarrow \langle \tilde{x}_j, \tilde{x}_k \rangle_{\mathcal{X}}$
  end for
  for $l = 1$ to $|\Sigma|$, where $\alpha^{(l)} \in \Sigma$ **do**
    $x^{(l)} \leftarrow (\alpha^{(l)} I + A^* A)^{-1} A^* y^e$; $\tilde{\kappa}_j^{(l)} \leftarrow \langle \tilde{x}_j, x^{(l)} \rangle_{\mathcal{X}}$
    if $l \geqslant 2$ **then**
      $\Delta^{(l)} \leftarrow \left| \tilde{\kappa}_j^{(l)} - \tilde{\kappa}_j^{(l-1)} \right|$
    end if
  end for
  $l_* = \arg\min\{\Delta^{(l)}, \ 2 \leqslant l \leqslant |\Sigma|\}$; $\tilde{\kappa}_j \leftarrow \tilde{\kappa}_j^{(l_*)}$
end for
$\tilde{\boldsymbol{\beta}} \leftarrow \boldsymbol{G}^{-1} \tilde{\boldsymbol{\kappa}}$, where $\boldsymbol{G} := (G_{j,k} : \ j, k \in \mathbf{N}_n)$
return $x_{ag} \leftarrow \tilde{\boldsymbol{x}} \tilde{\boldsymbol{\beta}}$

3.3 Using aggregation as a feasible solution to the joint inversion problem

With the above discussion, we are now ready to provide a landscape view of the proposed aggregation scheme for the resolvent of the joint inversion problem. Consider the joint inversion problem of multiple observation (1) and (2). We summarize below the general method applied to the joint inversion problem in Algorithm 3.2.

Next, we prove convergence of Algorithm 3.2.

Theorem 3.7. *If n vectors $\boldsymbol{\lambda}^{(j)} := \left(\lambda_i^{(j)}, \ i \in \mathbf{N}_m\right)$, $j \in \mathbf{N}_n$ of weighted parameters for the objective functional (5) are provided, $x^\dagger \in A_\varphi(R)$, where φ is an index function such that the function $\omega(t) = \frac{t}{\varphi(t)}$*

Algorithm 3.2 for joint inversion

Input: $y_i^{e_i}, A_i,\ i \in \mathbf{N}_m,\ \boldsymbol{\lambda}^{(j)} := \left(\lambda_i^{(j)} :\ i \in \mathbf{N}_m\right),\ j \in \mathbf{N}_n,\ \Sigma$

Output: $\tilde{x}$

1: $x^{\boldsymbol{e}}_{\boldsymbol{\lambda}^{(j)}} \leftarrow (I + \mathbb{A}^*_{\boldsymbol{\lambda}^{(j)}}\mathbb{A}_{\boldsymbol{\lambda}^{(j)}})^{-1}\mathbb{A}^*_{\boldsymbol{\lambda}^{(j)}}\boldsymbol{y}^{\boldsymbol{e}}$, for all $j \in \mathbf{N}_n$

2: $y^e, A \leftarrow$ **select** the most trustable ones from input $y_i^{e_i}$, and the corresponding operators $A_i :\ i \in \mathbf{N}_m$

3: **return** $\tilde{x} = x_{ag} \leftarrow$ **call** Algorithm 3.1 with parameters $A, y^e, \tilde{\boldsymbol{x}} := \left(\tilde{x}_j = x^{\boldsymbol{e}}_{\boldsymbol{\lambda}^{(j)}} :\ j \in \mathbf{N}_n\right), \Sigma$

is increasing, $\omega(0) = 0$, and A is the operator corresponding to the most trustable observation chosen from (1), then the aggregation scheme depending only on the given data generates the approximant $\tilde{x} = x_{ag}$ of $x^\dagger$ satisfying

$$\|x^\dagger - \tilde{x}\|_{\mathcal{X}} \leqslant \min_j \left\{\|x^\dagger - x^{\boldsymbol{e}}_{\boldsymbol{\lambda}^{(j)}}\|_{\mathcal{X}}\right\} + o(\varphi(\theta^{-1}(\varepsilon))),$$

where $x^{\boldsymbol{e}}_{\boldsymbol{\lambda}^{(j)}}$, $j \in \mathbf{N}_n$, are the minimizers of the objective functional Φ defined by (5) with the parameters $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(j)}$.

Proof: For each $\boldsymbol{\lambda}^{(j)},\ j \in \mathbf{N}_n$, let

$$\tilde{x}_j := x^{\boldsymbol{e}}_{\boldsymbol{\lambda}^{(j)}} = \left(I + \mathbb{A}^*_{\boldsymbol{\lambda}^{(j)}}\mathbb{A}_{\boldsymbol{\lambda}^{(j)}}\right)^{-1}\mathbb{A}^*_{\boldsymbol{\lambda}^{(j)}}\boldsymbol{y}^{\boldsymbol{e}}.$$

By using the results of Theorem 3.6 with all conditions fulfilled, we have that

$$\|x^\dagger - \tilde{x}\|_{\mathcal{X}} = \|x^\dagger - x_{ag}\|_{\mathcal{X}} = \|x^\dagger - x_*\|_{\mathcal{X}} + o(\varphi(\theta^{-1}(\varepsilon))),$$

and $\tilde{x} = x_{ag}$ can be effectively attained by applying the aggregation scheme only depending on the input data: multiple observation (1), (2) and $\boldsymbol{\lambda}^{(j)},\ j \in \mathbf{N}_n$. Moreover, by the definition of x_*, we obtain that

$$\|x^\dagger - x_*\|_{\mathcal{X}} \leqslant \min_j \left\{\|x^\dagger - \tilde{x}_j\|_{\mathcal{X}}\right\} = \min_j \left\{\|x^\dagger - x^{\boldsymbol{e}}_{\boldsymbol{\lambda}^{(j)}}\|_{\mathcal{X}}\right\}.$$

Combining the above discussion, we get the desired estimate. □

Theorem 3.7 reveals that when coping with a joint inversion problem while several possible weighted parameter vectors for observation errors are provided and the best is not decided, an "aggregation" approach may achieve a more reliable result. Such an "aggregation of weighted parameters" is interpreted in a way that it aggregates the solution candidates corresponding to each parameter vector. The "reliability" is in the sense that the error between the aggregator and the real solution will not exceed the one between the solution candidate corresponding to any single weighted parameter and the real solution, plus a term of higher order than the best guaranteed accuracy of the reconstruction error from the most trustable observation equation one believes. Further, the retrospection of Theorem 3.6 also supports that in general, better accuracy of the aggregated solution than any single solution may be expected. The above theoretical suggestions are justified by the numerical tests in the following section.

4 Numerical Illustrations

We present in this section numerical experiments to demonstrate the efficiency of the proposed aggregation method and to compare it with other existing methods in the literature. Our numerical experiments are preformed with MATLAB version 8.4.0.150421 (R2014b) on the PC CELSIUS R630 Processor Intel(R) Xeon(TM) CPU 2.80 GHz. All data are simulated in a way that they mimic the inputs of the SST-problem and the SGG-problem described by the equations (1) with (3).

It is well-known (see, e.g., [9]) that the integral operators A_i defined by (3) with the kernels $h_i,\ i = 1, 2$, act between the Hilbert spaces $\mathcal{X} = L_2(\Omega_{R_0})$ and $\mathcal{Y}_i = L_2(\Omega_{\rho_i})$ of square-summable functions on the spheres $\Omega_{R_0}, \Omega_{\rho_i},\ i = 1, 2$, and admit the singular value expansions

$$A_i x(\cdot) = \sum_{k=0}^{\infty} a_k^{(i)} \sum_{l=0}^{2k+1} \frac{1}{\rho_i} Y_{k,l}\left(\frac{\cdot}{\rho_i}\right) \left\langle \frac{1}{R_0} Y_{k,l}\left(\frac{\cdot}{R_0}\right), x(\cdot)\right\rangle_{L_2(\Omega_{R_0})}, \tag{32}$$

where $\{Y_{k,l}(\cdot)\}$ is the orthonormal system of the spherical harmonics on the unit sphere Ω_1, and

$$a_k^{(1)} = \left(\frac{R_0}{\rho_1}\right)^k \frac{k+1}{\rho_1},\ a_k^{(2)} = \left(\frac{R_0}{\rho_2}\right)^k \frac{(k+1)(k+2)}{\rho_2^2},\ k \in \mathbf{N}_0 := \{0,1,\ldots\}. \tag{33}$$

The solution $x = x^\dagger$ to (1) with (3) models the gravitational potential measured at the sphere Ω_{R_0}, that is expected to belong to the spherical Sobolev space $H_2^{\frac{3}{2}}(\Omega_{R_0})$ (see, e.g., [27]), which means that its Fourier coefficients $\left\langle \frac{1}{R_0} Y_{k,l}\left(\frac{\cdot}{R_0}\right), x^\dagger \right\rangle_{L_2(\Omega_{R_0})}$ should be at least of order $O((k+1)^{-\frac{3}{2}})$, $k \in \mathbf{N}_0$. Therefore, to produce the data for our numerical experiments we simulate the vectors

$$x_k^\dagger = \left\{ \left\langle \frac{1}{R_0} Y_{k,l}\left(\frac{\cdot}{R_0}\right), x^\dagger \right\rangle_{L_2(\Omega_{R_0})} : \ l \in \mathbf{N}_{2k+1} \right\},\ k \in \mathbf{N}_0,$$

of the Fourier coefficients of the solution $x^\dagger$ in the form

$$x_k^\dagger = (k+1)^{-\frac{3}{2}} \xi_k,\ k \in \mathbf{N}_0,$$

where ξ_k are random $(2k+1)$-dimensional vectors, whose components are uniformly distributed on $[-1,1]$. In view of (32), the vectors of the Fourier coefficients

$$y_{k,i}^{e_i} = \left\{ \left\langle \frac{1}{\rho_i} Y_{k,l}\left(\frac{\cdot}{\rho_i}\right), y_i^{e_i} \right\rangle_{L_2(\Omega_{\rho_i})} : \ l \in \mathbf{N}_{2k+1} \right\},\ k \in \mathbf{N}_0$$

of noisy data $y_i^{e_i}$ are simulated as

$$y_{k,i}^{e_i} = a_k^{(i)} x_k^\dagger + e_i,\ k \in \mathbf{N}_0,\ i = 1,2,$$

where $e_1, e_2 \in \mathbb{R}^{2k+1}$ are Gaussian white noise vectors which roughly correspond to (2) with $\varepsilon_1 : \varepsilon_2 = 3:1$. All random simulations are performed 500 times such that we have data for 1000 problems of the form (1) with (3). Moreover, we take $R_0 = 6371(km)$ for the radius of the Earth, and $\rho_1 = 6621(km), \rho_2 = 6771(km)$. All spherical Fourier coefficients are simulated up to the degree $k = 300$, which is in agreement with the dimension of the existing models, such as Earth Gravity Model 2008 (EGM2008). Thus, the set of simulated problems consists of 500 pairs of the SGG- and SST-type problems (1) with (3). In our experiments, each pair is inverted jointly by means of Tikhonov-Phillips regularization (4), (5) performed in a direct weighted sum of the observation spaces $\mathcal{Y}_i = L_2(\Omega_{\rho_i})$, $i = 1,2$, and we use three methods for choosing the regularization parameters (weights) λ_1, λ_2.

In the first method (i.e. M1), we relate them according to (7). Recall that the data are simulated such that $\varepsilon_1 : \varepsilon_2 = 3:1$. Therefore, we have $\lambda_2 = 9\lambda_1$. Then the parameter λ_1 is chosen according to the standard quasi-optimality criterion from the geometric sequence $\Sigma_{30} = \{\lambda^{(j)} = 10^{\frac{40+j}{8}} : j \in \mathbf{N}_{30}^0\}$. As a result, for each of 500 pairs of the simulated problems we apply Algorithm 2.1 and obtain a regularized approximation to the solution $x^\dagger$ that will play the role of the approximant $\tilde{x}_1$.

In the second method (i.e. M2), the parameters λ_1, λ_2 are selected from Σ_{30} according to the multiple version of the quasi-optimality criterion. In this way, for each of 500 pairs of the simulated problems we apply Algorithm 2.2 and obtain the second approximant $\tilde{x}_2$.

The third method (i.e. M3) consists in aggregating the approximnats $\tilde{x}_1, \tilde{x}_2$ according to the methodology described in the end of Subsection 3.2. In our experiments the role of the most trustable observation equation (14) is played by the equations of the SGG-type (3), (32), $i = 2$, because the data for them are simulated with smaller noise intensity. Then the required regularization parameters α_1, α_2 are selected from the geometric sequence Σ_{30} in such a way that $\frac{1}{\alpha_1}, \frac{1}{\alpha_2} \in \Sigma_{30}$ according to the quasi-optimality criterion (31). In this way, for each of 500 pairs of $\tilde{x}_1, \tilde{x}_2$, we apply Algorithm 3.1 and obtain an aggregated solution x_{ag}.

The performance of all three methods is compared in terms of the relative errors $\|x^\dagger - \tilde{x}_j\|_{\mathcal{X}} / \|x^\dagger\|_{\mathcal{X}}$, $j = 1,2$, and $\|x^\dagger - x_{ag}\|_{\mathcal{X}} / \|x^\dagger\|_{\mathcal{X}}$. The results are displayed in Figure 4.1, where the projection of each circle onto the horizontal axis exhibits a value of the corresponding relative error of one of the methods M1, M2, and M3 in a joint inversion of one of 500 pairs of the simulated problems. From this figure we can conclude that the aggregation by the linear functional strategy (Method M3) essentially improves the accuracy of joint inversion compared to the aggregated methods. This conclusion is in agreement with our Theorem 3.6.

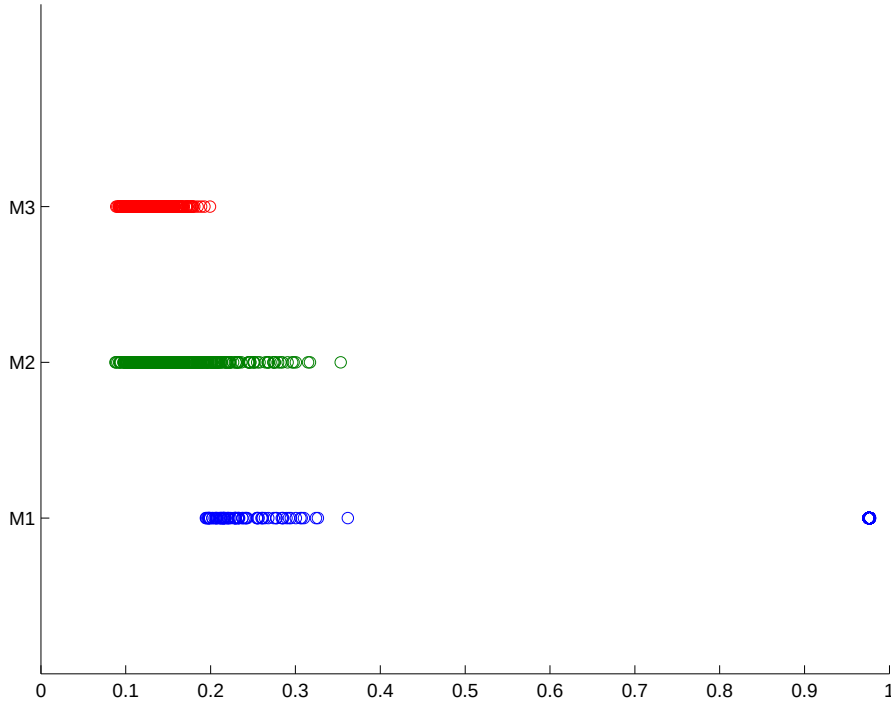

Figure 4.1: Examples of a joint regularization of two observation models. Relative errors of the regularization by a reduction to a single regularization parameter (M1), the regularization with a multiple quasi-optimality criterium (M2), and the regularization by aggregation (M3).

Note that it is also possible to aggregate the approximants $\tilde{x}_j$ corresponding to different values of the regularization parameter of a one parameter regularization scheme applied to a single observation equation (1). In such a case this equation is also used within the linear functional strategy to approximate the components $\kappa_j = \langle \tilde{x}_j, x^\dagger \rangle_{\mathcal{X}}$ of the vector $\boldsymbol{\kappa}$. This method is labeled with M4 and Algorithm 4.1 is provided below. To illustrate this method (i.e. M4) we consider the SST-type problem (3),(32), $i = 1$. We choose this problem because for the considered values $R_0 = 6371(km), \rho_1 = 6621(km), \rho_2 = 6771(km)$ it is less ill-posed than the SGG-type problem (3),(32), $i = 2$. This can be seen from (33) when one compares the rates of the decrease of the singular values $a_k^{(1)}$ and $a_k^{(2)}$ as $k \to \infty$: both decrease exponentially fast, but $a_k^{(1)}$ decreases slower than $a_k^{(2)}$.

When applying Algorithm 4.1, we use the data simulated as above and for each of 500 equations of the SST-type we construct the candidate approximants $\tilde{x}_j,\ j \in \mathbf{N}_{30}$ of the form

$$\tilde{x}_j = (\alpha_j I + A_1^* A_1)^{-1} A_1^* y_1^{e_1},$$

where α_j traverses the geometric sequence $\Sigma'_{30} = \{\alpha^{(j)} = 10^{-\frac{225+j}{30}} : j \in \mathbf{N}_{30}^0\}$. In one way, we choose the final approximant x_{Q_1} according to the standard quasi-optimality criterion [28], which we label as "Q1". Then, in the other way, we continue Algorithm 4.1 by performing aggregation upon the candidate approximations $\tilde{x}_j,\ j \in \mathbf{N}_{30}$ with the most trustable observation equation being selected by the SST-type problem itself and using the same geometric sequence Σ'_{30}.

Again the performance of the methods Q1 and M4 is compared in terms of the relative errors $\|x^\dagger - x_{Q_1}\|_{\mathcal{X}}/\|x^\dagger\|_{\mathcal{X}}$, and $\|x^\dagger - x_{ag}\|_{\mathcal{X}}/\|x^\dagger\|_{\mathcal{X}}$. The results are displayed in Figure 4.2, where the circles have the same meaning as in Figure 4.1. Figure 4.2 shows that the aggregation based on the linear functional strategy, which is equipped with the quasi-optimality criterion (31), essentially improves the accuracy of the regularization, as compared to the use of the standard quasi-optimality criterion.

It is also instructive to compare Figure 4.1 with Figure 4.2. This comparison shows that the aggregation of the approximants coming from joint inversion of SGG- and SST-type models outperforms the

Algorithm 4.1 for M4

Input: A, y^e, Σ' where Σ' is the parameter set

Output: x_{ag}

for $j' = 1$ to $|\Sigma'|$, where $\alpha^{(j')} \in \Sigma'$ **do**
  $\tilde{x}^{(j')} \leftarrow (\alpha^{(j')} I + A^* A)^{-1} A^* y^e$
  for $k = 1$ to j' **do**
    $G_{k,j'} = G_{j',k} \leftarrow \left\langle \tilde{x}^{(j')}, \tilde{x}^{(k)} \right\rangle_{\mathcal{X}}$
  end for
end for
for $j = 1$ to $|\Sigma'|$ **do**
  for $l = 1$ to $|\Sigma'|$ **do**
    $\tilde{\kappa}_j^{(l)} \leftarrow \left\langle \tilde{x}^{(j)}, \tilde{x}^{(l)} \right\rangle_{\mathcal{X}}$
    if $l \geqslant 2$ **then**
      $\Delta^{(l)} \leftarrow \left| \tilde{\kappa}_j^{(l)} - \tilde{\kappa}_j^{(l-1)} \right|$
    end if
  end for
  $l_* \leftarrow \arg\min\{\Delta^{(l)},\ 2 \leqslant l \leqslant |\Sigma'|\};\ \tilde{\kappa}_j \leftarrow \tilde{\kappa}_j^{(l_*)}$
end for
$\tilde{\boldsymbol{\beta}} \leftarrow \boldsymbol{G}^{-1} \tilde{\boldsymbol{\kappa}}$
return $x_{ag} \leftarrow \tilde{\boldsymbol{x}} \tilde{\boldsymbol{\beta}}$, where $\tilde{\boldsymbol{x}} = \left(\tilde{x}^{(j)} : j \in \mathbf{N}_{|\Sigma'|} \right)$

aggregation of the approximants coming from the single observation model. Clearly, Figures 4.1 and 4.2 demonstrate the advantage of the joint inversion.

5 Concluding Remarks

There is a need to reduce the uncertainty of biases when adopting regularization parameter choice methods, especially in the multi-parameter choice problem arising from the multi-penalty or multi-observation inversion, e.g., joint inversion of multiple observations. It is of no doubt that suitable parameter choices are critical for the accuracy of reconstruction which needs to be considered carefully in practice. Such a matter usually depends on a concrete problem at hand, the *a priori* information we know from the problem and the method we decide to use. However, unfortunately it is of no hope to gather the whole information usually. Therefore, parameter choice schemes in the light of *a priori* or *a posteriori* principle emerge but it seems that not a single one can avoid practice bias and drawback in certain extent, to the extent of our knowledge. Moreover, different solutions resulted from different parameter choice methods may compensate the flaw of each other. The methodology of aggregation is such a method to meet the above requirements that a more reliable result with more tolerance of the reconstruction error can be expected in a normal circumstance.

We conclude that the proposed linear aggregation of approximate solutions resulting from different methods to the same quantity of interest might allow significant improvement of the reconstruction accuracy and reduction of the uncertainty of possible bias by using a single method. In another angle of view, aggregation might be treated as a third method of the parameter choice scheme among, or an ideal supplement to, the classical and state-of-art parameter choice methods in the literature. But it still needs to point out that the efficiency of aggregation relies on the quality of "material" to be aggregated and the underlying methods. In this sense, it is reasonable that the extent of "good quality" and "bad quality" of the material should be within the tolerance of errors and the extent of co-linearity of the material for aggregation in order to expect a better aggregation which, of cause, might exhibit different effects in practice. Our theory guarantees the aggregation accuracy will not exceed any one by using only a single parameter choice method plus a term of order higher than the best guaranteed reconstruction accuracy from the most trustable observation equation one believes. Our experiments show that the results in

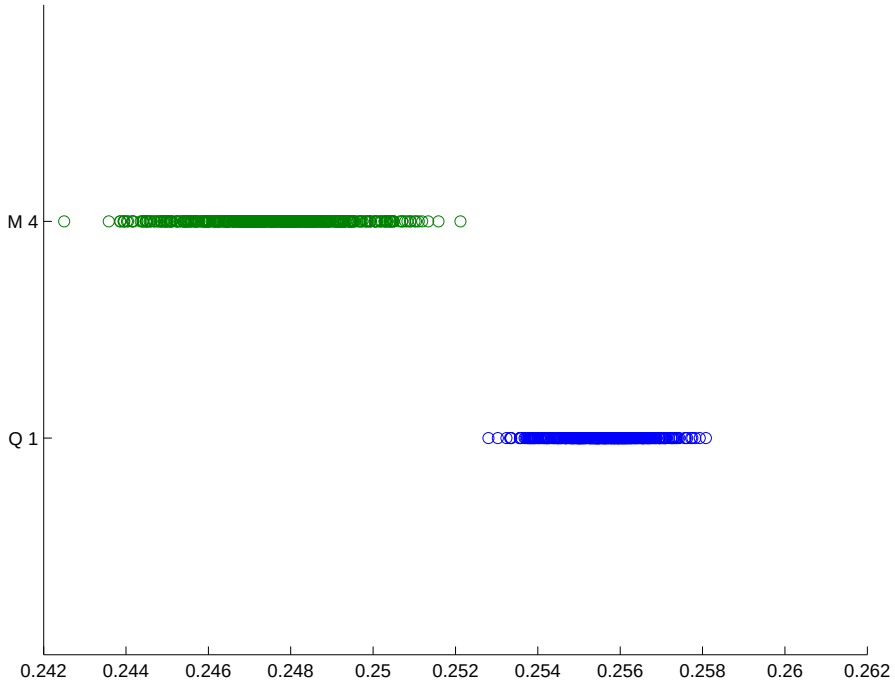

Figure 4.2: Examples of the regularization of a single observation model. Relative errors of the regularization with the quasi-optimality criterion (Q1) and of the regularization by aggregation (M4).

practice might be more promising in most general cases.

Acknowledgments

This research was performed when the first author visited Johann Radon Institute for Computational and Applied Mathematics (RICAM) within the framework of cooperation between RICAM and Guangdong Province Key Lab of Computational Science. The stay of the first author at RICAM was supported in part by the Austrian Science Fund (FWF), Grant P25424 and by Guangdong Provincial Government of China through the "Computational Science Innovative Research Team" program. The research of the third author was supported in part by the United States National Science Foundation under grant DMS-1115523, by the Guangdong Provincial Government of China through the Computational Science Innovative Research Team program and by the Natural Science Foundation of China under grants 11071286, 91130009 and 11471013.

References

[1] Anderssen R S 1986 The linear functional stragegy for improperly posed problems *Inverse Problems* Birkhaeuser, Basel 11–30

[2] Anderssen R S and Engl H W 1991 The role of linear functionals in improving convergence rates for parameter identification via Tikhonov regularization in: Yamaguti M, Hayakawa K, Iso Y, Mori M, Nishida T, Tomoeda K,Yamamoto M (eds.) *Inverse Problems in Engineering Sciences, ICM-90 Satellite Conf. Proc. Springer, Tokyo* 1–10

[3] Anderssen R S 2005 Inverse problems: A pragmatist's approach to the recovery of information from indirect measurements *ANZIAM Journal* 46 C588–C622

[4] Bailey D L, Hutton B F, and Walker P J 1987 Improved SPECT using simultaneous emission and transmission tomography *Journal of Nuclear Medicine : official publication, Society of Nuclear Medicine* 28 (5) 844–851

[5] Bauer F, Mathe P and Pereverzev S 2007 Local solutions to inverse problems in geodesy *Journal of Geodesy* 81 (1) 39–51

[6] Brezinski C, Redivo-Zaglia M, Rodriguez G, et al 2003 Multi-parameter regularization techniques for ill-conditioned linear systems *Numerische Mathematik* 94 (2) 203–228

[7] Brenner M, Jiang Y, and Xu Y 2008 Multiparameter regularization for Volterra kernel identification via multiscale collocation methods *Advances in Computational Mathematics* 31 (4) 421–455

[8] Chen Z, Lu Y, Xu Y, and Yang H 2008 Multi-parameter Tikhonov regularization for linear ill-posed operator equatioins *Jurnal of Computational Mathematics* 26 (1) 37–55

[9] Freeden W 1999 Multiscale modeling of spaceborne geodata *Teubner*

[10] Gerhards C 2013 A combination of downward continuation and local approximation for harmonic potentials *Inverse Problems 30 085004*

[11] Goldenshluger A and Pereverzev S V 2000 Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations *Probability Theory and Related Fields* 118 (2) 169–186

[12] Gyulai A, Baracza M K and Tolnai E E 2013 The application of joint inversion in geophysical exploration *International Journal of Geosciences* 4 283

[13] Hegland M 1992 An optimal order regularization method which does not use additional smoothness assumptions *SIAM Journal of Numerical Analysis* 29 (5) 1446–1461

[14] Ito K, Jin B, and Takeuchi T 2011 Multi-parameter Tikhonov regularization *arXiv preprint arXiv 1102.1173*

[15] Jones L K 1992 A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training *The Annals of Statistics* 608–613

[16] Juditsky A and Nemirovski A 2000 Functional aggregation for nonparametric regression *The Annals of Statistics* 681–712

[17] Koch K R and Kusche J 2002 Regularization of geopotential determination from satellite data by variance components *Journal of Geodesy* 76 (5) 259–268

[18] Kunisch K and Pock T 2013 A bilevel optimization approach for parameter learning in variational models *SIAM Journal on Imaging Sciences* 6 (2) 938–983

[19] Louis A K 1996 Approximate inverse for linear and some nonlinear problems *Inverse problems* 12 (2) 175–190

[20] Lu S and Pereverzev S V 2013 Regularization Theory for ill-posed problems - Selected Topics *de Gruyter, Berlin/Boston*

[21] Lu S, Pereverzyev Jr S and Sampath S 2013 Multiparameter Regularization for Construction of Extrapolating Estimators in Statistical Learning Theory *Multiscale Signal Analysis and Modeling* Chapter 15 Springer New York 347-366

[22] Lu Y, Shen L, and Xu Y 2007 Multi-parameter regularization methods for high-resolution image reconstruction with displacement errors *Circuits and Systems I: Regular Papers, IEEE Transactions on* 54 (8) 1788–1799

[23] Mathe P 2006 The Lepskii principle revisited *Inverse Problems* 22 (3) L11–L15

[24] Mathe P and Hofmann B 2008 How general are general source conditions? *Inverse Problems* 24 (1) 015009

[25] Mathe P and Pereverzev S V 2003 Geometry of linear ill-posed problems in variable Hilbert scales *Inverse Problems* 19 (3) 789

[26] Merrick J R W 2013 Aggregation of forecasts from multiple simulation models *C. Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World. IEEE Press* 533–542

[27] Svensson S L 1983 Pseudodifferential operators - a new approach to the boundary problems of physical geodesy *Manuscr. Geod.* 8 (1) 1–40

[28] Tikhonov A N and Glasko V B 1965 Use of the regularization method in non-linear problems *USSR Computational Mathematics and Mathematical Physics* 5 (3) 93–107

[29] Vozoff K and Jupp D L B 1975 Joint inversion of geophysical data *Geophysical Journal International* 42 (3) 977–991

[30] http://www.esa.int/Our_Activities/Observing_the_Earth/GOCE

[31] http://www.gfz-potsdam.de/champ/

[32] http://www.jpl.nasa.gov/missions/gravity-recovery-and-climate-experiment-grace/